PROTOCOL 3V2640-CLIN-005

A PHASE 2, MULTI-CENTER, SINGLE-BLIND, RANDOMIZED, PLACEBO CONTROLLED STUDY OF TVB 2640 (ASC40) IN SUBJECTS WITH NON-ALCOHOLIC STEATOHEPATITIS

Statistical Analysis Plan

FINAL VERSION 1
DATE OF PLAN:

02June2020

STUDY DRUG:
TVB-2640

PREPARED FOR:

Sagimet Biosciences

[redacted]

[redacted]

[redacted] Date

[redacted]

Approval Signature Page: Sagimet Biosciences

[redacted]

Date

[redacted]

Date

Contents

ABBREVIATIONS

Abbreviation	Explanation
15-HETE	15-LOX–derived 15-hydroxyeicosatetraenoic acid
Adipo-IR	Adipose tissue insulin resistance
AE	Adverse event
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
ApoB	Apolipoprotein B
AST	Aspartate aminotransferase
ATC	Anatomic therapeutic class
ATP	As-Treated Population
BMI	Body mass index
CI	Confidence interval
CK-18	Cytokeratin-18
cm	Centimeter
CMH	Cochran-Mantel-Haenszel
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
DM	Diabetes mellitus
DNL	De novo lipogenesis
ECG	Electrocardiogram
eCRF	Electronic case report form
ELF	Enhanced liver function
FASN	Fatty acid synthase
FDA	Food and Drug Association
FIB-4	Fibrosis-4
GGT	Gamma-glutamyl transpeptidase
HbA1c	Glycated hemoglobin
HDL-C	High-density lipoprotein cholesterol
HOMA-IR	Homeostatic model assessment insulin resistance
ICH	International Council for Harmonisation
IL-6	Interleukin 6
INR	International normalized ratio
ITT	Intent-to-treat population
IWRS	Interactive Web Response System
kg	Kilogram
LDL-C	Low-density lipoprotein cholesterol
LLN	Lower limit of normal
LOCF	Last Observation Carried Forward
Lp(a)	Lipoprotein(a)
LTB4	5-LOX–derived leukotriene B4
LXA4	Lipoxin A4
m	Meter
max	Maximum
MedDRA	Medical Dictionary for Regulatory Activities

Abbreviation	Explanation
min	Minimum
mITT	Modified intent-to-treat population
mmHg	Millimeters of mercury
MMRM	Mixed model for repeated measures
MRE	Magnetic resonance elastography
ms	Millisecond
N or n	Number of subjects
NASH	Non-alcoholic steatohepatitis
NCI	National Cancer Institute
NEFA	Non-esterified fatty acid
MRI-PDFF	Proton density fat fraction by magnetic resonance imaging
PGE2	prostaglandin E2
PK	Pharmacokinetic
PO	Orally
PP	Per-protocol
PT	Preferred term
Q1	First quartile
Q3	Third quartile
QD	Once daily
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SE	Standard error
SI	International System of Units
SOC	System organ class
SRC	Safety Review Committee
TEAE	Treatment-emergent adverse event
TLF	Tables, listings, and figures
TG	Triglycerides
TNFα	Tumor necrosis factor-alpha
ULN	Upper limit of normal
VCTE	Vibration-controlled transient elastography
β-HCG	Beta-human chorionic gonadotropin

1. INTRODUCTION

The statistical analysis plan (SAP) details the planned analysis required to satisfy the Clinical Study Report (CSR) of Sagimet Biosciences, Inc. (Sagimet) study number 3V2640-CLIN-005: A Phase 2, Multi-Center, Single-Blind, Randomized, Placebo-Controlled Study of TVB-2640 in Subjects With Non-Alcoholic Steatohepatitis (NASH). The content of this SAP is based on the protocol version 6.0 dated 29Apr2020.

SAP Revision Chronology:

Final Version 1	02June2020	Original

2. STUDY OBJECTIVES AND ENDPOINTS

2.1. Primary

The primary objectives of the study are:

- To determine the effect of once daily (QD) TVB-2640 for 12 weeks versus placebo on the change in hepatic fat fraction by proton density fat fraction by magnetic resonance imaging (MRI-PDFF) from baseline in subjects with NASH.
- To determine the safety of QD TVB-2640 versus placebo in subjects with NASH, including the effects on alanine aminotransferase (ALT) and aspartate aminotransferase (AST).

2.2. Secondary

The secondary objective is:

- To determine the effect of QD TVB-2640 for 12 weeks versus placebo in subjects with NASH on:
  - Lipid and lipoprotein parameters, including low-density lipoprotein cholesterol (LDL-C), non-LDL-C, high-density lipoprotein cholesterol (HDL-C), non-HDL-C, total cholesterol, triglycerides (TG), apolipoprotein B (ApoB), and lipoprotein(a) (Lp[a]) particles.
  - NASH and fibrosis biomarkers including cytokeratin-18 (CK-18), fibrosis-4 (FIB-4), and enhanced liver function (ELF) or FIBROSpect 2 or ProC3 test.
  - Eicosanoid panel.

2.3. Exploratory

The exploratory objectives are:

- To determine the effect of QD TVB-2640 for 12 weeks versus placebo in subjects NASH on clinical measures, including:
  - Liver fibrosis, as determined by vibration-controlled transient elastography (VCTE).

  - Metabolic parameters, including fasting glucose, insulin, homeostatic model assessment insulin resistance (HOMA2-IR), glycated hemoglobin (HbA1c), non-esterified fatty acid (NEFA), adipose tissue insulin resistance (adipo-IR), adiponectin, resistin, Interleukin 6 (IL-6), and gamma-glutamyl transpeptidase (GGT).
  - Anthropometric parameters, including weight, waist and hip circumference, waist-hip ratio, and blood pressure.
  - Lipidomic analyses for de novo lipogenesis (DNL).
- Explore the relationship of plasma drug exposure to changes in efficacy and safety biomarkers.
- Explore possible relationships between genomic markers of NASH and responses to treatment.

3. STUDY DESIGN

3.1. Study Design and Population

This is a multi-center, randomized, single-blind, placebo-controlled, dose-escalation study to evaluate the safety and efficacy of TVB-2640 in subjects with NASH. Male and female subjects aged ≥18 years with either biopsy-proven NASH within 2 years before randomization and MRI-PDFF ≥8% or, if prior liver biopsy was not performed or results are not available, magnetic resonance elastography (MRE) ≥2.5 kpa and MRI-PDFF ≥8% during screening are planned to be enrolled. Approximately 117 subjects will be enrolled and randomized, with at least 90 subjects enrolled in the US and approximately 27 subjects enrolled in China (18 active and 9 placebo). In the US, subjects will be enrolled and randomized to achieve 90 evaluable subjects (30 active and 15 placebo in each of 2 dose cohorts). The 50 mg cohort will enroll and randomize approximately 27 additional subjects from China. A subject is defined as evaluable if they receive study drug for at least 8 weeks and have a baseline and at least 1 post-baseline MRI-PDFF assessment on or after Week 8.

Subjects will receive TVB-2640 or placebo QD orally (PO) for 12 weeks. During the 12-week Treatment period, subjects are to attend study center visits on Days 1 and 2, and Weeks 2, 4, 8, and 12. After completion of the 12-week Treatment period, subjects are to attend a Follow-up visit at Week 16 for post-treatment safety and efficacy assessments.

All subjects must complete the 12-week treatment period at the 25 mg dose level, with review of all safety data and written approval by the Safety Review Committee (SRC) and reviewed by the Food and Drug Administration (FDA) before enrollment and treatment of subjects at the 50 mg dose level may commence.

Due to the impact of COVID-19, several subjects were unable to have their Week 12 end of treatment visit performed within the protocol defined window. These subjects received extended dosing of study product and will have their end of treatment visit collected more than 12 weeks post first dose of study product, with allowable ranges between Week 11 and Week 16. Study

drug was not permitted to be continued for more than 16 weeks. For these subjects, the End of Study Visit was scheduled for 4 weeks after the last dose of study drug. For purposes of consistent terminology in context of the COVID changes, the following terminology will be used,

End of Treatment Visit: Visit performed following the end of study treatment. This will be the Week 12 visit for subjects who completed the treatment period as planned. For subjects who entered extended dosing, the visit will be performed on or prior to Week 16. For subjects who received less than 12 weeks of study treatment, the visit will be prior to the scheduled Week 12 visit.

End of Study Visit: Visit corresponding to 4 weeks after the end of treatment visit. This will be the Week 16 visit for subjects who completed the treatment period as planned.

3.2. Randomization and Blinding

Subjects who are candidates for enrollment into the study will be assigned a sequential and unique subject number by the Investigator after the subject has provided written informed consent. Once a subject number has been assigned, it cannot be reused.

Subjects who have provided written informed consent will be evaluated for eligibility by the Investigator to ensure that the entry criteria (see Protocol Sections 8.2 and 8.3) have been satisfied and that the subjects is eligible for participation in this clinical study.

Two TVB-2640 dose levels are planned to be evaluated in a sequential fashion: 25 mg and, if study suspension criteria are not met at the 25 mg dose level (see Protocol Section 7.2), 50 mg. At each dose level, subjects will be randomly assigned to TVB-2640 or placebo at a 2:1 ratio, with randomization stratified by type 2 diabetes mellitus (DM) status. The 50 mg dose level will also be stratified by country:

- 25 mg dose level: TVB-2640 (Planned N=30) and Placebo (Planned N=15)
- 50 mg dose level: US: TVB-2640 (Planned N=30) and Placebo (Planned N=15); China: TVB-2640 (Planned N=18) and Placebo (Planned N=9); Total: TVB-2640 (Planned N=48) and Placebo (Planned N=24)

Randomization will be done via Interactive Web Response System (IWRS). Subjects are to receive their first study drug dose on Day 1 within 24 hours after randomization.

At both dose levels, subjects will receive TVB-2640 or placebo QD PO for 12 weeks.

Study drug will be dispensed to subjects in a single-blinded fashion. All other study personnel, with the exception of personnel conducting/interpreting imaging studies (see Protocol Section 11.1) at each study center, will be unblinded to the subject's treatment assignment.

3.3. Sample Size Considerations

The sample size is based on a fixed-sequence strategy which tests for a treatment difference in the primary endpoint (percent change at Week 12 in liver fat from baseline as determined by

MRI-PDFF) between each TVB-2640 dose group versus placebo. The fixed-sequence strategy will start with the highest dose group comparison. If statistically significant, then testing will proceed to the low dose group comparison. Each TVB-2640 dose group will be compared to the pooled placebo group using an F-test test from an analysis of covariance (ANCOVA) model with fixed effects for the stratification factor (diabetes presence/absence) and treatment group (i.e., TVB-2640 dose groups and pooled placebo) and with the baseline MRI-PDFF value as a covariate. Since the primary efficacy endpoint may not be normally distributed, the non-parametric Wilcoxon rank-sum test is conservatively used, instead of the F-test, for power calculations.

The sample size is based on a conservative assumption that the primary analysis is going to be performed in the US population only, resulting in 30 evaluable subjects in the placebo group, and 30 evaluable subjects in each of the TVB-2640 dose levels. Based on Patel (2016), it is conservatively assumed that the primary endpoint has a standard deviation of 30. Power calculations assume the primary endpoint is lognormally distributed with a standard deviation of 30, there are at least 30 evaluable subjects in each treatment group, and the two-sided Wilcoxon rank sum test will be used to test each pairwise treatment difference at the 0.05 Type I error level. Under the fixed-sequence strategy which maintains an overall 0.05 Type I error rate, the study has at least 80% overall power to detect both treatment differences (i.e., both high dose versus placebo and low dose versus placebo), if each mean treatment difference is at least 24. If the study does not proceed to the high-dose level, (i.e., there are 30 evaluable subjects treated with 25 mg TVB-2640 and only 15 with placebo), then the study has at least 77% power to detect a mean treatment difference of at least 24. The study will have more power if the subjects from China are included in the primary analysis and the same assumptions of treatment difference and variability hold.

Power calculations for the key secondary endpoint (percentage of subjects with at least a 30% reduction in the primary endpoint) are based on pairwise comparisons with the placebo group using Fisher's two-sided exact test [equivalent to Cochran-Mantel-Haenszel (CMH) without stratification] at the 0.05 significance level. Based on Loomba (2018), it is expected there will be less than 5 responses out of 30 evaluable subjects in the placebo group (i.e., less than 16.7%). With 30 evaluable subjects in each treatment group, there is at least 80% power to detect at least a 36.7% increase in the response rate (TVB-2640 treatment group minus placebo) when the placebo response rate is at most 16.7%. If the study does not proceed to the high-dose level, (i.e., 30 evaluable subjects treated with 25 mg TVB-2640 and only 15 with placebo), then the study has at least 63% power to detect at least a 36.7% increase in the response rate (25 mg TVB-2640 treatment group minus placebo) when the placebo response rate is at most 2 out of 15 evaluable subjects (i.e., 13.34%).With 30 subjects in each TVB-2640 treatment arm, there is at least a 95% probability of detecting at least 1 AE when the AE rate is 10% or more in the underlying population.

3.4. Safety Review Committee

An SRC will oversee the study to ensure subject safety and to advise if any dosing alterations are recommended. The SRC will review safety including liver-related events (i.e., clinically

meaningful elevations in ALT, AST, and bilirubin), and other efficacy (lipid parameters) and safety data as needed.

If study suspension criteria are met at any time (see Protocol Section 7.2), the SRC will review the relevant study data and make a determination whether enrollment and treatment of subjects may continue or be permanently terminated. In addition, the SRC will review all safety data from the 25 mg dose level after all subjects at that dose level complete the 12-week treatment period. Enrollment and treatment of subjects at the 50 mg dose level may commence only with the written approval of the SRC and with FDA review. The SRC may also review data from the 50 mg cohort, if performed, upon request (see below).

Additional details on the SRC can be found in the SRC Plan. The SRC Plan includes details on membership, purpose, organization, data requirements, logistics, and reporting requirements.

3.5. Timing of Analyses

An interim analysis for SRC review will be performed after all of the subjects in the 25 mg dose level complete the 12-week treatment period. Details on data requirements and outputs required will be maintained in the SRC Plan or other related documents.

An additional SRC analysis was performed during the COVID pandemic to confirm safety prior to allowing the option for extended 50 mg cohort dosing.

The primary analysis of efficacy will be performed when all subjects in the US 50 mg cohort finish their final visit. The primary analysis will not include any data from China, due to delays and uncertainty of enrollment in the Chinese cohort.

An End of Study final analysis will be performed when the final subject completes the study and will include cumulative data from the US and China as applicable.

4. STUDY ENDPOINTS

4.1. Efficacy

The primary efficacy endpoint is:

- Percent change from baseline in liver fat at Week 12, as determined by MRI-PDFF.

The key secondary efficacy endpoint is:

- The percentage of subjects with at least a 30% reduction in liver fat at Week 12, as determined by MRI-PDFF.

Additional secondary efficacy endpoints are:

- Percent change from baseline in liver fat at End of Treatment and End of Study, as determined by MRI-PDFF.

- Percentage of subjects with at least a 30% reduction in liver fat at End of Treatment and End of Study, as determined by MRI-PDFF.
- Change from baseline over the treatment period in:
    - Liver aminotransferases.
    - Lipid and lipoprotein parameters.
    - NASH and fibrosis markers.
    - Eicosanoids.

Exploratory efficacy endpoints are:

- Change from baseline over the treatment period in:
    - Liver fibrosis, as determined by VCTE.
    - Metabolic parameters.
    - Anthropometric parameters.
    - Lipidomic analyses for DNL.
- Proportion of subjects with <5% liver fat (normalized) at Week 12.

4.2. Safety

Safety endpoints are:

- Proportion of subjects experiencing treatment-emergent adverse events (TEAEs), Grade 3
- or 4 TEAEs, and serious adverse events (SAEs).
- Change from baseline in vital sign measurements, 12-lead ECG findings, and clinical laboratory test results.
- Proportion of subjects experiencing treatment-emergent ophthalmologic abnormalities.

5. ANALYSIS POPULATIONS/SETS

The modified intent-to-treat population (mITT) is defined as all subjects who are randomized, receive study drug for at least 8 weeks, and have a baseline and at least 1 post-baseline MRI-PDFF assessment on or after Week 8. The mITT is the primary population for analysis of MRI-PDFF assessments and subjects will be analyzed according to the randomized treatment assignment.

The intent-to-treat population (ITT) is defined as all subjects who are randomized and received at least 1 dose of study drug. The ITT will be used for analysis of a subset of secondary efficacy endpoints and supportive and sensitivity analyses of MRI-PDFF assessments. Subjects will be analyzed according to the randomized treatment assignment.

The as-treated population (ATP) is defined as all subjects who are randomized and received at least 1 dose of study drug. The ATP will be used for analysis of secondary efficacy endpoints and supportive and sensitivity analyses of MRI-PDFF assessments. Subjects will be analyzed according to the treatment received.

The safety population is defined as all subjects who are randomized and received at least 1 dose of study drug and will be used for all analysis of safety. Subjects will be analyzed according to

the treatment received. If all subjects were dosed according to randomized treatment assignment, then the safety population and ITT are identical.

A per-protocol (PP) population may be defined for sensitivity analyses of the efficacy endpoints. The PP population will be based on subjects in the mITT population with data excluded from subjects who have important protocol violations deemed to possibly affect the efficacy measurements or integrity of any other key study data. Subjects will be analyzed according to the actual treatment received. If a PP population is implemented, details on subjects excluded will be provided in the CSR.

6. DEFINITIONS

7. GENERAL DATA HANDLING CONSIDERATIONS

All analyses will be conducted using SAS® software Version 9.4 or higher.

All data in the database will be presented in by-subject data listings.

Unless otherwise stated, all listings will be sorted by treatment group, center ID, subject number, and assessment date and time, if available.

Unless stated otherwise, continuous data will be summarized by treatment group using descriptive statistics including number of subjects (n), mean, median, standard deviation (SD), first quartile (Q1), third quartile (Q3), minimum (min) value, and maximum (max) value.

Unless stated otherwise, categorical data will be summarized by treatment group using counts and percentage based on the number of non-missing values. For selected summaries, the number of missing values will be presented as a separate category with no percentage, if one or more subjects have missing data. Counts of zero will be presented without percentages.

Descriptive statistics will be presented with the following numerical precision:

- Minimum and Maximum: same as the raw data collected
- Mean, Median, Q1, and Q3: one additional decimal place to that reported for Minimum and Maximum
- SD: two additional decimal places than the Minimum and Maximum
- Percentages: reported to one decimal place if percent is <100. If the value is 100, no decimal place will be reported.

Unless otherwise noted, statistical inference will be based on 2-sided testing with a 5% significance level (i.e. 95% confidence intervals [CI] will be produced).

P-values will be reported to four decimal places. If the value is below 0.0001 it will be noted as < 0.0001; if the value above 0.9999 it will be noted as > 0.9999.

All data up to the time of study completion/withdrawal from the study will be included in the analysis, regardless of duration of treatment.

Numbering for data displays will be based on International Council for Harmonisation (ICH) E3 whenever possible.

7.1. Stratification and Covariates

At each dose level, subjects will be randomly assigned to TVB-2640 or placebo at a 2:1 ratio, with randomization stratified by type 2 diabetes mellitus status (present/absent). The 50 mg dose level will also be stratified by country (US/China) for the final analysis. Selected statistical models and tests will be stratified based on the randomization strata. Details on the models and tests are provided in the relevant sections below.

This is a multi-center trial. However, adjustment for study center will not be made in the statistical tests.

7.2. Evaluation of Subgroups

All analyses will be performed separately by country (US and China) in addition to overall.

Selected analyses of liver fat reduction may also be repeated based on the following subgroups:

- ALT: Baseline ≥1.5 x ULN vs <1.5 x ULN
- DM status: present vs absent
- Baseline MRI-PDFF: ≥ median vs < median
- Baseline MRE: ≥ median vs < median
- Liver stiffness based on fibrosis stage: Stage 1 vs Stage 2/3
- Baseline triglycerides: ≥ median vs < median
- Weight change: >=5% loss vs <5% loss (or a gain)

7.3. Multiple Comparisons and Multiplicity

In order to control for multiple comparisons, the primary and key secondary efficacy analyses will be tested using a fixed-sequence strategy to maintain the overall Type I error rate at 0.05. Each test in the fixed-sequence, uses a 2-sided test at the 0.05 level of significance and starts with the primary efficacy analyses. The first comparison to be assessed for significance is the primary efficacy endpoint analysis of the high-dose group (50 mg TVB-2640) compared to placebo and if statistically significant, testing will proceed to the low-dose group (25 mg TVB-2640) comparison. If both comparisons of the primary efficacy endpoint are significant, then testing will proceed to the key secondary efficacy endpoint. The key secondary efficacy analyses will be tested in a similar fashion, first comparing the high dose group (50 mg TVB-2640) with placebo and if statistically significant, testing will proceed to the low dose group (25 mg TVB-2640) comparison.

7.4. Reference Dates

- Screening date is defined as the electronic case report form (eCRF) provided date on which a subject was screened for study entry.

- Randomization date is defined as the date on which the subject is randomized to study treatment.
- Treatment start date is defined as the date of first dose of study drug (also referred to as Day 1).
- Treatment end date is defined as the date of last dose of study drug.
- The calculation of age will use the screening date as its reference date.
- Safety data, such as AEs and laboratory assessments will use the treatment start date as a reference date to assign treatment emergence.
- Study day will be based on treatment start date as a reference date.

Reference day calculations will be defined as:

- date of interest – reference date + 1 when the date of interest >= reference date;
- otherwise, date of interest – reference date.

For example, duration of treatment is defined as treatment end date – treatment start date + 1.

If either date is missing, reference date calculations will not be performed. Date imputation will be performed as identified in Section 7.6.

7.5. Baseline and Post-Baseline Changes

Unless stated otherwise, baseline will be based on the last non-missing value collected prior to the start of study treatment based on the treatment start date and time, if applicable. Post-baseline values will be those collected after the treatment start date and time, if applicable.

Change from baseline is defined as:

post-baseline value – baseline value.

Percentage change from baseline is defined as:

(post-baseline value – baseline value)/baseline value X 100%.

7.6. Imputation of Partial Dates

Imputed dates will be used to classify treatment emergent AEs (TEAEs) and prior and concomitant medications.

Adverse Events

- If the AE start date is completely missing the event will be assumed to have started on the first day of treatment and will be considered a TEAE. If a subject was not treated, then no imputation will be conducted for missing start dates.
- If the AE start date is missing day and month, do the following:
    - If the AE start year is the same year as that of the first treatment and the AE contains information to indicate that the event ended before the treatment start

date (e.g. AE end date month and year are earlier than the treatment start date or the full AE end date is known and occurs earlier than the treatment start date), then set the AE day/month to '1 January'.
    - If the AE start year is the same as that of the first treatment and the AE contains information to indicate that the event ended after the treatment start date, then set the AE start day/month to the treatment start day/month
    - If the AE start year is later than that of the first treatment, then set the AE start day/month to '1 January.'
    - If the AE start year falls before that of the first treatment, then the AE start day/month will not be imputed.
- If only the AE start day is missing, do the following:
    - If the AE start year is the same year as that of the first treatment and the AE month is prior to the first treatment date, then set the AE day to '1'.
    - If the AE start month/year are the same as that of the first treatment and the AE contains information to indicate that the event ended before the treatment start date (e.g. the full AE end date is known and occurs earlier than the treatment start date), then set the AE day to '1'.
    - If the AE start month/year are the same as the first treatment and the AE end date is after the treatment start date, then set the AE day to the same day as treatment start day.
    - If the AE start year is the same as that of the first treatment and the AE start month is after the first treatment, or the AE start year is later than that of the first treatment, then set AE day to '1'.
- AE end dates will not be imputed.

Prior and Concomitant Medications

- The imputation rules for AE start dates will be used for prior and concomitant medication start dates.
- Prior and concomitant medication stop dates will be imputed as follows:
    - If the stop date is only missing the day, then the stop day is the last day of the month
    - If the stop date is missing both the day and month, then the stop month and day is December 31
    - If the stop date is completely missing, no imputation is performed, and the medication will be classified as a concomitant medication for subjects who were treated.

7.7. Multiple Assessments and Visit Windows

Nominal visits (e.g. those identified by the study eCRF) will be the basis of summarization and statistical analysis; no visit date windowing will be conducted. Unless specified otherwise, data from unscheduled visits will be included in: (1) summaries of most extreme and baseline data, (2) summaries of specific abnormalities at any time post-baseline, (3) adverse events and any other summaries not based on study visit, and (4) subject data listings.

7.8. Missing Data

Imputation will only be performed on the primary and selected secondary efficacy endpoints using MRI-PDFF data.

In the primary and key secondary analyses, missing Week 12 values for MRI-PDFF will be imputed with the last post-baseline observation on or after Week 8. For sensitivity analyses that include subjects with extended dosing, the End of Treatment values for MRI-PDFF will be imputed with the last post-baseline observation on or after Week 8.

MRI-PDFF analyses at other timepoints (e.g., End of Study) will follow similar conventions and use the last-post baseline observation on or after Week 8. Missing MRI-PDFF data before Week 8 will not be imputed.

Sensitivity analyses will be performed excluding data from subjects with missing data (i.e., observed cases approach).

Additional details are provided in the relevant analysis sections below.

8. ANALYSIS METHODS

8.1. Study Subject Data

8.1.1. Subject Disposition

Summaries of analysis population membership and final subject status (completed or withdrawn), including reasons for withdrawal, will be produced based on the number of randomized subjects. The number of subjects screened will be provided. Data will be presented by treatment group as well as overall for the study, with exception of the number of screen failures (to be displayed without respect to treatment).

Screen failures, analysis populations, and final subject disposition status will be listed.

8.1.2. Protocol Deviations

Protocol deviations will be identified and classified as important or not important. Important protocol deviations may include but are not limited to:

- Violation of Inclusion/Exclusion Criteria
- Study drug compliance ≤ 80% or ≥ 120%
- Use of prohibited therapies
- Incorrect treatment

Important protocol deviations deemed to possibly affect the efficacy measurements or integrity of any other key study data may be identified prior to database lock to support defining the PP population.

Protocol deviations including category will be summarized by treatment group and overall for the study. Protocol deviations due to COVID impact will be identified as such.

A listing of protocol deviations will be provided.

8.1.3. Demographic and Baseline Characteristics

Subject demographics will be summarized and listed. Age, sex (Male / Female), ethnicity (Hispanic or Latino / Not Hispanic or Latino), race (American Indian or Alaska Native / Asian / Black or African American / Native Hawaiian or Other Pacific Islander / White / Other), baseline height (cm), baseline weight (kg), body mass index (BMI) (kg/m^2), Type 2 Diabetes Mellitus status (Yes / No), waist circumference (cm), hip circumference (cm), waste-hip ratio (cm) and country (US / China). Age will also be categorized as a categorical variable (< 65, $65 \leq age < 75$, $75 \leq age < 85$, ≥ 85) for reporting. Data will be presented by treatment group as well as overall for the study.

The following conversions and equations will be used as applicable:

BMI (kg/m^2) = weight(kg)/[height$(m)^2$]

Demographic and baseline characteristics summaries will be provided for the mITT, and safety populations. Subgroup analyses will be produced by region (US/China).

Baseline characteristics including social history (tobacco, alcohol, and caffeine use) will be listed. Pregnancy test, breath/blood alcohol test, and beta-human chorionic gonadotropin (β-HCG) data will be listed. NASH confirmation data collected at screening will be presented in a data listing.

8.1.4. Medical History

Medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) version 22.0 Mar 2019 for reporting by system organ class (SOC) and preferred term (PT). Medical history data will be presented in a listing.

8.1.5. Prior and Concomitant Medication

The incidence of medication use will be summarized by WHO Drug Dictionary anatomic therapeutic class (ATC) Level 2 classification (i.e. therapeutic main group) and preferred name. A subject will be counted only once at each level of reporting. Prior medications are medications that have been discontinued prior to the treatment start date (e.g. taken exclusively during the pre-therapy period). Concomitant medications are medications taken at any point during the on-therapy or post-therapy periods. Subjects could start a concomitant medication prior to study start which is ongoing at time of first treatment or subjects could start a new medication during the on-therapy or post-therapy periods.

Concomitant medication use will be summarized separately and presented by treatment group for the ITT population.

All prior and concomitant medication data will be listed including the verbatim and preferred drug name and ATC Level 2.

8.2. Study Drug Exposure and Compliance

Duration of drug exposure will be summarized both as a continuous and a categorical variable for the safety population. Exposure will be calculated as:

study drug stop date – study drug start date + 1 [converted to weeks (days divided by 7)]

Categories for reporting exposure will include < 2 weeks; 2 weeks - <4 weeks; 4 weeks - <8 weeks; 8 weeks - <12 weeks; ≥12 weeks.

Drug accountability is captured in the eCRF at each visit. The eCRF captures the number of pills dispensed, the number returned, and the number missed during the dosing period. Overall compliance will be determined based on pill counts over each visit period and taking into account the number missed. Compliance will be derived as:

$$\frac{[(\#\text{ of pills dispensed} - \#\text{ of pills returned}) - \#\text{ of missed pills}] \times 100}{\text{number of days within the visit period}}$$

Where number of days within the visit period is derived as:

$$\text{Date of Visit (X+1)} - \text{Date of Visit (X)} + 1$$

Compliance will also be summarized for the following categories < 80%, 80% - 120%, and > 120%.

Mean daily dose (mg) and total cumulative dose (mg) received will also be summarized. Total cumulative dose will be summarized overall and separately for those that had extended dosing.

Listings of planned and actual treatments, overall compliance, and drug exposure will be produced.

8.3. Efficacy

8.3.1. Primary Efficacy Endpoint and Analyses

The primary efficacy endpoint is the percent change from baseline in liver fat at Week 12, as determined by MRI-PDFF. In the primary analysis, missing Week 12 values for MRI-PDFF will be imputed with the last post-baseline observation on or after Week 8 in the mITT (i.e., last observation carried forward [LOCF]).

8.3.1.1. Primary Analysis of the Primary Efficacy Endpoint

The initial analysis of the primary efficacy endpoint will be performed using an ANCOVA model with fixed effects for the stratification factor (diabetes presence/absence), and treatment group (TVB-2640 50 mg and pooled placebo)with the baseline MRI-PDFF value as a covariate including all subjects in the mITT from the US.

The primary efficacy analyses will be based on a F-test from the ANCOVA model in the mITT to compare each TVB-2640 dose group individually with the placebo group. Data normality will be assessed and the appropriate data transformation will be applied to fulfill the underlying assumption of the ANCOVA model. Model residuals will be calculated and plotted to visually inspect the model assumptions of homoscedasticity and normality for the primary efficacy analysis. If, upon visual inspection, the model assumptions of homoscedasticity or normality are not met, then the primary efficacy analysis will be repeated after making an appropriate transformation (e.g., log transformation) of the data. If a suitable transformation cannot be determined the corresponding rank ANCOVA will become the primary analysis for this endpoint.

The equal slopes assumption for the primary analysis will be assessed by including a term for the treatment by baseline MRI-PDFF interaction. If the equal slope assumption is not met (p-value for interaction less than 0.05), then the primary analysis will be repeated with the interaction term included in the model and the LS Means and treatment difference in the LS Means will be tested and summarized at the 25th, 50th, and 75th percentile for the baseline MRI-PDFF. This analysis will be considered secondary and supportive of the primary analysis.

Summary statistics will be displayed by treatment group along with the difference in least squares means (and the associated 95% confidence interval and p-value) for each TVB-2640 dose group to pooled placebo comparison. If data is transformed for analysis, then LS means will be back transformed to the original scale where appropriate.

The fixed-sequence method, as described in Section 7.3, will be used to maintain an overall 0.05 Type I error rate for the final analysis.

In the primary analysis, missing Week 12 values for MRI-PDFF will be imputed with the last post-baseline observation on or after Week 8 in the mITT.

Subjects who required extended dosing beyond 12 weeks will be excluded from the primary efficacy analysis. See Appendix 1 for additional details on COVID-19 impacts and data handling considerations.

The primary analysis will be based on the mITT population.

8.3.1.2. Additional Analyses of the Primary Efficacy Endpoint

The primary efficacy analysis will be repeated using the ITT population, ATP, and PP population. Analyses will be conducted with LOCF from Week 8 onward and repeated excluding missing data (no LOCF).

In addition to the primary analysis of Week 12 that includes all subjects who had no more than 12 weeks of study treatment, an analysis of the End of Treatment MRI-PDFF data including data from the extended dosing subjects will be performed in the mITT, ITT, ATP, and PP populations.

An analysis of MRI-PDFF data in the log scale will be performed in the mITT population. The log ratio of Week 12 to baseline [log(Week 12/Baseline)] will be analyzed using a similar

ANCOVA mode. In the event that model assumptions are not met for the primary efficacy analysis noted above, the analysis in the log scale will become primary.

At the time of the final analysis, the above analyses will be repeated using all data from the US and China. Subset analyses will be performed on the primary efficacy endpoints for subjects enrolled in US and China to evaluate the consistency of results between subjects enrolled in each country. The ANCOVA model will incorporate fixed effects for the stratification factor (diabetes presence/absence), country (US/China), treatment group (TVB-2640 50 mg and pooled placebo), and country-by-treatment interaction and with the baseline MRI-PDFF value as a covariate including all subjects in the mITT from US and China who received 50 mg or placebo.

A test for country-by-treatment interaction will be performed. If there is no interaction detected ($p \geq 0.05$), the primary efficacy endpoint, percent change from baseline in hepatic fat fraction at Week 12, as determined by MRI PDFF, will be analyzed using an ANCOVA model with fixed effects for the stratification factor (diabetes presence/absence) and treatment group (i.e., TVB-2640 dose groups and pooled placebo) and with the baseline MRI-PDFF value as a covariate including all subjects from US and China (with no fixed effect for country). If the country-by-treatment interaction is significant ($p<0.05$) then a separate analysis will be performed on subjects enrolled in China in the 50 mg cohort. This analysis is not powered to detect a significant difference but would be supportive/descriptive in nature to evaluate trends.

Analyses of absolute reduction from baseline to week 12 in liver fat will be performed in the mITT, ITT, ATP, and PP populations. Similar methodology as used for the primary efficacy analysis will be used.

In order to determine if the effect on liver fat persists after study treatment, the absolute and percent change in liver fat from End of Treatment to End of Study will be performed in the mITT population with missing data excluded. The analysis will be run with and without dosing extenders. Methodology similar to that described for the primary efficacy analysis will be used.

8.3.2. Secondary Efficacy Endpoints and Analyses

MRI-PDFF Endpoints

Week 12 data for the secondary efficacy endpoints will be handled similarly to the handling used for the primary efficacy endpoint. That is, analyses will be performed both excluding data from subjects who required extended dosing beyond 12 weeks and including data from the extended dosing subjects as supplemental analyses.

The key secondary efficacy endpoint is the percentage of subjects with at least a 30% reduction in liver fat at Week 12, as determined by MRI-PDFF. Note that due to data collection precision, subjects with a reduction of at least 29.5% from baseline will count as responders. Differences between each TVB-2640 dose group (25 mg, 50 mg) and pooled placebo for the key secondary efficacy endpoint will be analyzed using CMH methods. The key secondary efficacy analyses will be based on a CMH test in the mITT population to compare each TVB-2640 dose group with placebo, adjusting for the stratification factor (diabetes presence/absence). For the final end

of study analysis, if a country-by-treatment interaction is found in the primary efficacy endpoint, the subjects from the US and China in the 50 mg cohort will be analyzed separately for the key secondary efficacy endpoint.

To check for other possible interactions, i.e., diabetes strata-dependent treatment effects, the Breslow-Day test will be performed; if the p-value of the Breslow-Day test is ≥0.05, the treatment-by-strata interaction is deemed not significant. However, if the interaction effect is statistically significant (i.e., p value <0.05), then the Gail and Simon test will be used to test for the qualitative interaction at a significance level of 0.05 and provided as an aid for interpretation of the CMH results. The estimated response rates and the corresponding exact 95% confidence interval based on a binomial distribution will be calculated for each treatment group. The exact 95% confidence limit for the CMH estimate of the common risk difference will also be obtained; a Wald asymptotic approximation with continuity correction is an acceptable substitute for the exact limits in the event of computational difficulties.

Similar responder analyses will be performed using response level definitions of 20%, 40%, and 50%. As done for the key secondary endpoint efficacy endpoint, subjects with a reduction of at least 19.5%, 39.5%, and 49.5% from baseline, respectively, will count as responders. Analyses will be repeated in the ITT, ATP, and PP populations. Analyses of the key secondary endpoint performed in the ATP populations will include summaries of all subjects in the analysis population, as well as the subset of subjects that have at least 8 weeks of treatment and an evaluable assessment on/after 8 weeks (i.e., the same subjects in the mITT population).

Similar responder analyses will be performed using reduction in liver fat at End of Study in the mITT, ITT, and ATP populations.

A scatter plot of End of Treatment body weight changes versus End of Treatment liver fat changes will graphed by treatment and overall in the ATP. A subset will also be run in subjects with at least a 30% reduction in liver fat at Week 12.

In the key secondary analysis, missing Week 12 values for MRI-PDFF will be imputed with the last post-baseline observation on or after Week 8 in the mITT population. Sensitivity analyses excluding subjects with missing data will also be performed.

Other Efficacy Endpoints

Other efficacy endpoints will be summarized using the ATP. A limited set of endpoints will be repeated in the ITT.

Liver aminotransferases will be evaluated using observed values and changes from baseline by study visit. Summaries will be performed for ALT and AST and will be presented by treatment group.

The changes at 12 weeks will be summarized for each treatment group and compared between each TVB-2640 dose group and pooled placebo using a linear mixed-effects model for repeated measures (MMRM). The model will include the stratification factor (diabetes presence/absence),

treatment group, visit, and treatment-by-visit interaction as the fixed effects, and baseline value as a covariate, and the model will use an unstructured variance-covariance matrix. The point estimates for the least-squares mean of the pairwise difference for each TVB-2640 dose group and pooled placebo comparison at each visit and the corresponding 95% confidence interval and two-sided p-value will be summarized. If convergence problems arise when fitting the model, a spatial power covariance structure will be used as a first step. If the model still has convergence issues, compound symmetry covariance structure will be employed.

For the end of study final analysis, if a country-by-treatment interaction is found in the primary efficacy endpoint, the subjects from the US and China in the 50 mg cohort will be analyzed separately. However, if no country-by-treatment interaction is found in the primary efficacy endpoint, the interaction may be explored within the secondary endpoints in a similar manner in order to determine if there is a significant interaction for any of the other endpoints (e.g., ALT or AST).

The model produces asymptotically consistent estimates when data are missing at random. No imputation of missing data will be required for these analyses. By definition, subjects without a baseline and at least 1 post-baseline value will be excluded from the MMRM analyses. Model assumptions (normality of residuals and parallel slopes) may be assessed. If assumptions are not met, then transformations will be done analogous to the primary efficacy analyses described above.

Similar analyses will be performed for:

- Lipid and lipoprotein parameters including low-density lipoprotein cholesterol (LDL-C), non-LDL-C, high-density lipoprotein cholesterol (HDL-C), non-HDL-C, total cholesterol, TG, apolipoprotein B (ApoB), and lipoprotein(a) (Lp[a]) particles, total Cholesterol:HDL ratio, LDL:HDL ratio (Note that the cholesterol and LDL to HDL ratios will be derived and do not come directly from the laboratory
- NASH and fibrosis biomarkers including cytokeratin-18 (CK-18), fibrosis-4 (FIB-4), and enhanced liver function (ELF) or FIBROSpect 2 or ProC3 test.
- Eicosanoids including 5-LOX–derived leukotriene B4 (LTB4), COX-derived prostaglandin E2 (PGE2), and 15-LOX–derived 15-hydroxyeicosatetraenoic acid (15-HETE) and lipoxin A4 (LXA4). For eicosanoids, descriptive statistics will be presented without any statistical testing.

The number and percentage of subjects with normalized ALT will be summarized by study visit and treatment group. Differences between each TVB-2640 dose group (25 mg, 50 mg) and pooled placebo for ALT normalization will be analyzed using CMH methods. The analyses will be based on a CMH test to compare each TVB-2640 dose group with placebo, adjusting for the stratification factor (diabetes presence/absence). Only subjects with ALT>ULN at baseline will be included in the ALT normalization analyses.

Similar analyses will be performed for AST.

A scatter plot of End of Treatment TG changes versus End of Treatment liver fat changes will graphed by treatment and overall in the ATP. A subset will also be run in subjects with at least a 30% reduction in liver fat at Week 12

Subset analyses will be performed on relevant secondary efficacy endpoints for subjects enrolled in the US and China to evaluate the consistency of results between subjects enrolled in each country. The specific analyses performed by region will depend on the data that are ultimately collected in China. In the event that some efficacy parameters are not able to be evaluated in China, those analyses will not be performed.

8.3.3. Exploratory Efficacy Endpoints and Analyses

Analysis of exploratory endpoints will be based on the ATP. Exploratory efficacy endpoints include the change from baseline over the treatment period in following continuous measurements:

- Liver fibrosis, as determined by VCTE
- Metabolic parameters
- Anthropometric parameters including weight, waist and hip circumference, waist-hip ratio, and blood pressure
- Lipidomic analyses for DNL

Analyses will be performed similarly to those described in the secondary efficacy analysis section above.

The proportion of subjects with <5% liver fat (normalized) at Week 12 will be summarized by treatment group. Analyses similar to those described for the key secondary efficacy endpoint will be used.

Subset analyses will be performed on all exploratory efficacy endpoints for subjects enrolled in the US and China to evaluate the consistency of results between subjects enrolled in each country.

8.4. Pharmacokinetics (PK)

Blood samples for determination of TVB-2640 PK are to be collected immediately before and then at 2, 4, 6, and 24 hours after the first study drug dose; the 24-hour post-dose sample is to be collected on Day 2, immediately before the second study drug dose. Blood samples for PK also are to be collected immediately pre-dose and at 2, 4, and 6 hours post-dose at Week 2 (i.e., Day 8).

PK samples will be collected from 12 active subjects/cohort in the US and 8 active subjects in the 50 mg cohort in China.

PK analyses will be performed by a separate vendor and described outside of this SAP (i.e., in a separate PK SAP or other document).

8.5. Safety

Safety will be evaluated based on data collected for adverse events, vital signs, physical examinations, electrocardiogram (ECG), ophthalmologic examinations, and laboratory data. No

formal statistical testing will be conducted for the safety analyses. Descriptive statistics will be used to evaluate safety data in this study. Summaries will be presented by treatment group.

Subset analyses will be performed on all safety endpoints for subjects enrolled in China to evaluate the consistency of results between subjects enrolled in each country.

All safety analysis reporting will be based on the Safety Analysis Set. All safety data will be listed by treatment group and subject.

8.5.1. Adverse Events

An AE is any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product, and which does not necessarily have to have a causal relationship with this treatment. An AE can therefore be any unfavorable and unintended sign (including abnormal laboratory findings), symptom, or disease temporally associated with the use of an investigational product, whether or not related to the investigational product.

An unexpected AE is any event for which the nature or severity is not consistent with the information in the current Investigator's Brochure.

Any abnormal clinical or laboratory finding considered by the Investigator to be clinically significant is to be recorded in the eCRF as part of the subject's medical history if occurring prior to the start of study drug administration and as an AE if occurring after the start of study drug administration at baseline, where the finding represents a change from baseline.

A treatment emergent AE (TEAE) is defined as any AE occurring after the first dose of study treatment. All summaries of AEs will be based on TEAEs. All AEs recorded on the eCRF, whether treatment emergent or not, will be presented in the data listing. Non-TEAEs will be flagged in the listings.

Adverse Events will be coded using MedDRA version 22.0 Mar 2019 for reporting by SOC and PT. Summaries of TEAEs by SOC and PT will be presented in descending order of overall incidence. TEAEs will also be summarized by maximum intensity as classified according to the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE), Version 5.0. If the AE is not included in the NCI CTCAE, then the Investigator is to determine the severity of the AE according to the criteria provided in the study protocol. The causal relationship of each AE to study drug will be determined by the Investigator according to best medical judgment as definitely related, probably related, possibly related, unlikely related, or unrelated. If the relationship between the AE and study drug is determined to be "possible", "probable", or "definite", the event will be considered to be treatment-related for the safety analyses.

Any missing severity assessments will be assumed to be Grade 3 (severe) and missing relationship assessments will be assumed to be related. All AE summaries will be produced using the safety population and displayed by treatment group and overall.

An overview of TEAEs will be produced, including counts and percentages of subjects with any incidences of: TEAEs, TEAEs related to study treatment, serious adverse events (SAEs), TEAEs leading to study drug discontinuation, SAEs, and fatal SAEs. The number of AEs will also be included on the summary.

In addition to the overview of TEAEs, the following summaries of AEs will be provided for the safety population by treatment group and overall:

- TEAEs by SOC and PT
- TEAEs related to study treatment by SOC and PT
- CTCAE Grade 3 or higher TEAEs by SOC and PT
- CTCAE Grade 3 or higher TEAEs related to study treatment by SOC and PT
- TEAEs leading to treatment discontinuation by SOC and PT

Summaries of TEAEs by SOC, PT, and maximum severity as well as TEAEs by SOC, PT, and maximum relationship to study treatment will also be prepared.

A comprehensive listing of all AEs will be provided in a by-subject data listing. AEs that began after 84 days of dosing will be identified on the listing. In addition, the following listings will be provided:

- TEAEs related to study treatment
- SAEs
- TEAEs leading to treatment discontinuation; and
- Fatal AEs.

8.5.2. Clinical Laboratory Evaluations

Clinical chemistry, hematology, coagulation, and urinalysis parameters will be reported based on conventional units as applicable. The following laboratory evaluations will be reported in data summaries:

- Chemistry: Chloride, Carbon dioxide, Sodium, Potassium, Blood urea nitrogen, Calcium, Creatinine, Magnesium, Albumin, Glucose, Total protein, Alkaline phosphatase, AST, ALT, Indirect and direct bilirubin, Creatine phosphokinase (total and fractionated) (screening and baseline only). Note that AST and ALT collected as part of the clinical chemistry panel are also applicable for efficacy assessments.
- Hematology: Hematocrit, Hemoglobin, Red blood cell count, Platelet count, White blood cell count with differential, Absolute neutrophil count.
- Coagulation Studies: Prothrombin time, Activated partial thromboplastin time, international normalized ratio (INR).
- Urinalysis: Specific gravity, Protein, pH, Ketones, Blood, Microscopic examination of sediment, Glucose.

Observed values and changes from baseline for laboratory evaluations will be summarized at each visit and will include the most extreme change from baseline. The percent change from

baseline at each visit and the most extreme percent change from baseline will also be summarized. 95% confidence intervals for the change and percent change will be included.

Laboratory data will also be summarized in shift tables of baseline to each visit and most extreme change based on range categories of low (below lower limit of normal [LLN], normal, and high [above upper limit of normal [ULN]). Urinalysis results and shifts will be based on the categories present in the data.

8.6. Other Safety Evaluations

8.6.1. Vital Signs

Vital signs include: respiratory rate (breaths/minute); temperature (^{o}C); and pulse rate (beats/minute). Note that systolic and diastolic blood pressure will be summarized with the other anthropometric parameters. Observed values and changes from baseline for vital signs will be summarized at each visit and time point, as well as for most extreme change from baseline. Summaries will be provided for the safety population and displayed by treatment group.

8.6.2. Electrocardiogram (ECG)

Electrocardiogram (ECG) parameters include: HR (bpm), PR (ms), QRS (ms), QT (ms), QTcF (ms), QTcB (ms), and RR interval (ms). Observed values and changes from baseline for ECG parameters will be summarized by treatment group at each visit and time point.

Investigator reported ECG result will be tabulated by visit and worst-case post-baseline. In addition, the shifts from baseline to each visit and worst-case post-baseline will be summarized. Worst case post-baseline will be based on the worst observed value on or after the treatment start date. Categories will include within normal limits, abnormal without clinical significance, and abnormal with clinical significance.

8.6.3. Ophthalmologic Examinations

The proportion of subjects experiencing ophthalmologic abnormalities will be tabulated by visit and by treatment group. The clinical interpretation of the ophthalmologic examination will be tabulated as normal or abnormal. The worst-case post-baseline will also be summarized.

8.6.4. Physical Examinations

Physical examination results will be presented in subject data listings.

9. CHANGES TO THE PLANNED ANALYSES

The timing of the primary efficacy analysis will be based on completion of the US 50 mg cohort. The protocol indicated the initial analysis of the primary efficacy endpoint would be performed using subjects from US and China and include country as an effect in the ANCOVA model. Given the delays in enrollment in China due to COVID-19, the primary analysis has been adjusted to only include the US cohorts and a modified ANCOVA model as described above.

The treatment group used for the presentation of data based on the mITT population has been modified from the protocol. Per the study protocol, the data were to be analyzed based on randomized treatment.

As a point of clarification and as noted for the key secondary efficacy analysis above, due to data collection precision, subjects with a reduction of at least 29.5% from baseline will count as responders. Similar conventions were used for the sensitivity analyses of that endpoint using different responder definitions.

Since one subject received the incorrect study treatment (active instead of placebo), an as-treated population (ATP) has been added to the SAP. The ATP will be used to evaluate the treatment differences in efficacy endpoints using treatment received instead of randomized treatment.

Due to the impact of COVID-19, all efficacy endpoints will be analyzed at End of Treatment and End of Study. The definition of those timepoints are described in section 3.1. Per the protocol, endpoints were evaluated at Week 12 and Week 16 but the differences in duration of treatment require further evaluation given the visits may not correspond to the number of weeks specified in the protocol visit schedule.

An additional SRC analysis was performed during the COVID-19 pandemic to confirm safety prior to allowing the option of extended 50mg cohort dosing. If additional SRC reviews are performed, they will be discussed in the CSR.

Additional supplementary and sensitivity analyses related to the protocol-planned analyses have been added to the SAP. Specifically, the absolute change from baseline in MRI-PDFF has been added in addition to the planned percent change. The change from Week 12 to Week 16 in MRI-PDFF has been added. ALT and AST normalization has been added as an additional analysis of the liver aminotransferases.

Analyses and data presentations described in the protocol have been expanded upon, but no other changes were made to analyses specified in the protocol. Any additional changes to the planned analyses as detailed in this SAP will be described in the CSR.

APPENDIX 1. COVID-19 IMPACT AND DATA HANDLING

COVID-19 impacts on 3V2640-CLIN-005 include, but are not limited to, missed visits, inability to complete certain assessments per the protocol including Week 12 MRI-PDFF data which constitutes the data for the primary efficacy endpoint. In order to collect as much efficacy data on all randomized subjects, subjects were allowed to continue on their treatment beyond Week 12 and have their end of treatment MRI-PDFF performed at any time between Week 11 and Week 16 per protocol Amendment 5. Study drug was not permitted to be continued for more than 16 weeks. The End of Study Visit was scheduled for 4 weeks after the last dose of study drug.

Given the uncertainty of the impact of extended dosing beyond Week 12 on efficacy, the primary and secondary efficacy analyses will be performed excluding subjects who entered extended dosing beyond 12 weeks. Sensitivity analyses that include these subjects will be performed to evaluate the end of treatment efficacy.

The analysis of safety will include all collected data regardless of treatment extension.

An addendum to this SAP will be developed that documents all additional data handling techniques employed to manage COVID impacted data, if needed.

APPENDIX 2. PLANNED TABLES, LISTINGS, AND FIGURES

The table of contents and mock tables, listings, and figures (TLF) are maintained in a separate document.

PROTOCOL 3V2640-CLIN-005

A PHASE 2, MULTI-CENTER, SINGLE-BLIND, RANDOMIZED, PLACEBO CONTROLLED STUDY OF TVB-2640 (ASC40) IN SUBJECTS WITH NONALCOHOLIC STEATOHEPATITIS

Statistical Analysis Plan Addendum

VERSION 1
DATE OF PLAN:

14Mar2022

STUDY DRUG:
TVB-2640

PREPARED FOR:

Sagimet Biosciences

[redacted]

Approval Signature Page: [redacted]

[redacted]	Date

Approval Signature Page: Sagimet Biosciences

[redacted]

Date

Contents

1. INTRODUCTION

This document serves as an addendum to the statistical analysis plan (SAP) version 1 dated 02June2020. The SAP Addendum is focused on details pertaining to the planned analyses for the TVB-2640 75mg Open-Label Cohort 3 and should be read along with the primary study SAP that detailed the primary analyses of data from Cohort 1 and Cohort 2.

The content of this SAP Addendum is based on the protocol version 7.0 dated 17Mar2021.

SAP Revision Chronology:		
Final Version 1	02June2020	Original
Addendum 1	14Mar2022	Addendum to Original SAP

2. DESCRIPTION OF ANALYSES FOR COHORT 3

Data collected for the thirteen subjects in safety Cohort 3 will be summarized similarly to data collected for the first two study cohorts. Data collected via the EDC (CRF data) and from external vendors (laboratory data and MRI-PDFF) will be summarized, graphically presented, and listed. All summaries will be based on descriptive statistics. Continuous data will be summarized using descriptive statistics including number of subjects (n), mean, median, standard deviation (SD), first quartile (Q1), third quartile (Q3), minimum (min) value, and maximum (max) value.

Categorical data will be summarized using counts and percentage based on the number of non-missing values. For selected summaries, the number of missing values will be presented as a separate category with no percentage, if one or more subjects have missing data. Counts of zero will be presented without percentages.

Statistical significance will not be assessed for this study cohort. If any p-values are generated, they will be evaluated in order to assess study trends and not to assess statistical significance.

Statistical outputs will follow the design of those prepared for prior analyses but will include only a single column for the Cohort 3 (75mg) dose group.

Details on the analysis results and any changes from the protocol and/or prior SAP will be described in the clinical study report (CSR).

APPENDIX 2. PLANNED TABLES, LISTINGS, AND FIGURES

Type	Number	Title	Population
Table	14.1.1	Summary of Subject Disposition	All Subjects
Table	14.1.2.1	Summary of Subject Demographics	Modified Intent-to-Treat Population
Table	14.1.2.2	Summary of Subject Demographics	Safety Population
Table	14.1.3.1	Summary of Subject Baseline Characteristics	Modified Intent-to-Treat Population
Table	14.1.3.3	Summary of Subject Baseline Characteristics	Safety Population
Table	14.1.4.1	Summary of NASH History	Modified Intent-to-Treat Population
Table	14.1.4.3	Summary of NASH History	Safety Population
Table	14.1.5	Summary of Protocol Deviations	Intent-to-Treat Population
Table	14.1.6	Summary of Concomitant Medications	Safety Population
Table	14.1.7	Summary of Study Drug Exposure and Compliance	Safety Population
Table	14.2.1.1.1	Summary of Liver Fat (%) Based on MRI-PDFF	Modified Intent-to-Treat Population
Table	14.2.1.7	Summary of Liver Fat (%) Based on MRI-PDFF Log Ratio Versus Baseline	Modified Intent-to-Treat Population
Table	14.2.2.1.1	Summary of Liver Fat (%) Based on MRI-PDFF - Responder Analysis	Modified Intent-to-Treat Population
Table	14.2.3.1.1	Summary of Liver Aminotransferase Results and Change from Baseline	As-Treated Population
Table	14.2.3.4	Summary of Percentage of ALT and AST Normalization by Visit: Missing Data Excluded	As-Treated Population
Table	14.2.4.1.1	Summary of Lipid and Lipoprotein Results and Change from Baseline	As-Treated Population
Table	14.2.7	Summary of Liver Fibrosis Results as Determined by VCTE	As-Treated Population
Table	14.2.8.1.1	Summary of Metabolic Parameter Results and Change from Baseline	As-Treated Population
Table	14.2.9	Summary of Anthropometric Body Measurement Parameter Results and Change from Baseline	As-Treated Population
Table	14.3.1	Overall Summary of Adverse Events	Safety Population
Table	14.3.2	Summary of Treatment Emergent Adverse Events by SOC and PT	Safety Population
Table	14.3.3	Summary of Treatment Related Treatment Emergent Adverse Events by SOC and PT	Safety Population
Table	14.3.5	Summary of Treatment Emergent Adverse Events Leading to Treatment Discontinuation of Study Treatment by SOC and PT	Safety Population
Table	14.3.7	Summary of Treatment Emergent Adverse Events by SOC, PT, and CTCAE Grade	Safety Population
Table	14.3.8	Summary of Treatment Related Treatment Emergent Adverse Events by SOC, PT, and CTCAE Grade	Safety Population
Table	14.3.9	Summary of Treatment Emergent Adverse Events by SOC, PT, and Relationship to Treatment	Safety Population
Table	14.3.10	Summary of Selected Clinical Laboratory Results and Changes from Baseline	Safety Population
Table	14.3.14	Summary of Clinical Chemistry Shifts from Baseline	Safety Population
Table	14.3.18	Summary of Vital Signs Results and Changes from Baseline	Safety Population

Type	Number	Title	Population
Table	14.3.19	Summary of ECG Results and Changes from Baseline	Safety Population
Table	14.3.20	Summary of Clinically Significant Ophthalmologic Examination Abnormalities	Safety Population
Listing	16.2.1.1	Patient Disposition	All Subjects
Listing	16.2.1.2	Informed Consent and Randomization	All Subjects
Listing	16.2.2	Protocol Deviations	All Subjects
Listing	16.2.4.1	Demographics	All Subjects
Listing	16.2.4.2	Social History	All Subjects
Listing	16.2.4.3	Baseline Internal Laboratory Results	All Subjects
Listing	16.2.4.4	Medical History	All Subjects
Listing	16.2.4.5	NASH Confirmation Results	All Subjects
Listing	16.2.4.6	Physical Examination Results	All Subjects
Listing	16.2.4.7	Prior and Concomitant Medications	All Subjects
Listing	16.2.5.1	Study Drug Accountability	All Subjects
Listing	16.2.5.2	Study Drug Administration	All Subjects
Listing	16.2.5.3	Study Drug Dosing Adjustments	All Subjects
Listing	16.2.6.1	MRI-PDFF Results	All Subjects
Listing	16.2.6.2	Lipid and Lipoprotein Parameter Results	All Subjects
Listing	16.2.7.1	All Adverse Events	All Subjects
Listing	16.2.7.2	Treatment Related Treatment Emergent Adverse Events	All Subjects
Listing	16.2.7.3	Serious Adverse Events	All Subjects
Listing	16.2.7.4	Treatment Emergent Adverse Events Leading to Treatment Discontinuation	All Subjects
Listing	16.2.7.5	Treatment Emergent Adverse Events of Special Interest	All Subjects
Listing	16.2.7.6	Fatal Adverse Events	All Subjects
Listing	16.2.8	Selected Laboratory Results	All Subjects
Listing	16.2.8.1	Clinical Chemistry Results	All Subjects
Listing	16.2.8.2	Hematology Results	All Subjects
Listing	16.2.8.3	Coagulation and Urinalysis Results	All Subjects
Listing	16.2.8.4	Other Laboratory Results	All Subjects
Listing	16.2.8.5	Vital Signs	All Subjects
Listing	16.2.8.6	ECG Results	All Subjects
Listing	16.2.8.7	Ophthalmologic Examination Results	All Subjects
Listing	16.2.8.8	Anthropometric Body Measurements	All Subjects
Listing	16.2.8.9	FibroScan Results	All Subjects
-	-	-	-
Figure	14.2.1.2.1	Mean (SE) Percent Change from Baseline in Liver Fat (%) Based on MRI-PDFF	Modified Intent-to-Treat Population

Type	Number	Title	Population
Figure	14.2.1.3	Mean (SE) Change from Baseline in Liver Fat (%) Based on MRI-PDFF	Modified Intent-to-Treat Population
Figure	14.2.2.5	Plot of MRI-PDFF Responder Percentage Week 12 LOCF	Modified Intent-to-Treat Population
Figure	14.2.3.2.1	Mean (SE) ALT Over Time	As-Treated Population
Figure	14.2.3.2.2	Mean (SE) ALT Over Time	As-Treated Population with Baseline, Week 12 and Week 16 Data
Figure	14.2.3.5	Percentage of Subjects with ALT Normalization Over Time	As-Treated Population with Baseline ALT>ULN
Figure	14.2.3.3	Mean (SE) AST (U/L) Over Time	As-Treated Population
Figure	14.2.3.6	Percentage of Subjects with AST Normalization Over Time	As-Treated Population with Baseline AST>ULN
Figure	14.2.4.2	Mean (SE) Triglycerides (mg/dL) Over Time	As-Treated Population
Figure	14.2.4.3	Box and Whisker Plot of Triglycerides (mg/dL) Over Time	As-Treated Population
Figure	14.2.4.4	Scatter Plot of End of Treatment Triglyceride (mg/dL) Changes versus Liver Fat Absolute Changes (%)	As-Treated Population
Figure	14.2.4.5	Scatter Plot of End of Treatment Triglyceride (mg/dL) Changes versus Liver Fat Absolute Changes (%)	As-Treated Population with At Least a 30% Reduction in Liver Fat at End of Treatment
Figure	14.2.4.6.1	Scatter Plot of End of Treatment ALT Changes (%) versus Liver Fat Absolute Changes (%)	As-Treated Population
Figure	14.2.4.6.2	Scatter Plot of End of Treatment ALT Changes (%) versus Liver Fat Percentage Changes	As-Treated Population
Figure	14.2.4.7	Scatter Plot of End of Treatment ALT Changes (%) versus Liver Fat Changes (%)	As-Treated Population with At Least a 30% Reduction in Liver Fat at End of Treatment
Figure	14.2.4.8	Scatter Plot of End of Treatment Weight (kg) Changes versus Liver Fat Changes (%)	As-Treated Population
Figure	14.2.4.9	Scatter Plot of End of Treatment Weight (kg) Changes versus Liver Fat Changes (%)	As-Treated Population with At Least a 30% Reduction in Liver Fat at End of Treatment

Output numbers and titles may vary slightly in the final outputs. Any significant changes from the main study SAP to the Cohort 3 analyses will be described in the CSR.